NEARFIELD SYNTHESIS OF COMPLEX SOURCES WITH HIGH-ORDER AMBISONICS, AND BINAURAL RENDERING

Dylan Menzies

Dept. Computer Science and Engineering
De Montfort University
Leicester UK
dylan@dmu.ac.uk

ABSTRACT

A scheme is presented for encoding general complex sources in high-order Ambisonic soundfields, with control over position and orientation. Also reviewed is related work by the author on the binaural rendering of nearfield sources, accounting fully for the physical constraints of this problem. Together these developments provide a means for creating high quality nearfield auditory displays over headphones.

1. INTRODUCTION

A natural goal in many auditory displays is to be able to reproduce the perception of an arbitrary audio environment as well as possible. In addition to this it is often desirable to be able to compose the environment from audio objects, which can be located and transformed in flexible ways. This paper addresses the problem of composing an environment given by a high-order Ambisonic description, [1, 2], using source objects that can each model any finite source region of sound to any desired accuracy. By its generality this scheme includes the composition of near objects which convey a complex soundfield to the listener. Such soundfields provide valuable information about source location, and it is also suggested that they can help greatly in terms of display realism due to the complex nature of the near soundfields. For example, we might wish to create the perception of listening to a violin 50cm away, and be able to move and rotate the violin. The high-order Ambisonic encoding can be used to render a display over loudspeakers, however we focus here on rendering by headphones, otherwise called binaural rendering, as this is likely to be the most practical method for high definition nearfield displays. Speaker systems are limited by the number of speakers and the acoustics within which they operate. Binaural systems have no theoretical limitations, but do require good quality headtracking in order for the sound presented to be consistent with the head direction. Headtracking systems have improved rapidly in recent years both in quality and cost.

We begin by reviewing recent work [3] on the binaural rendering of nearfield monopole sources. This exposes a problem with rendering such soundfields, for which a solution is offered. Following on from this, a method is found for transforming an encoding of a general source to a listener-based soundfield encoding, taking into account the position and orientation of the source. Figure 1 is a very simple schematic of the process being considered. The soundfield encoding can be applied directly to the previously described binaural rendering, and more generally with speaker array rendering systems.

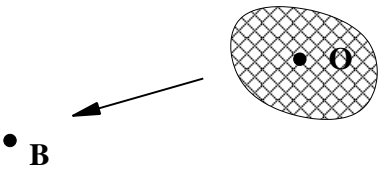

Figure 1: Encoding a complex source. O denotes the extended source object, and B the central point of the listener.

2. BINAURAL RENDERING OF SOUNDFIELD ENCODINGS

2.1. Ambisonic encoding

Ambisonics is a methodology developed for encoding soundfields, and decoding them onto speaker arrays, [4, 5]. Initially it was used only to 1st order, with 4 signals that encode a full sphere of sound around a central listener. More recently Ambisonics has been employed at higher orders, whereby it is possible to not only increase the angular resolution of distant sources, but also extend the listening region and recreate accurately the soundfield from nearfield sources, [2]. We shall refer informally to an encoding of any order as *B-format*, borrowing the original terminology for 1st order. Using high-order encodings, the listener receives distance cues about near sources exactly as they would for the real soundfield, because the soundfield around the listener can be reconstructed arbitrarily well.

The Ambisonic encoded signals are defined by a spherical harmonic expansion of a source-free region of the pressure field, also called the interior expansion because it is valid for general fields within a radius that contains no source region. Although our discussion does not depend on a particular representation, for definiteness we use signals, $B^{\sigma}_{mn}(k)$, defined with the real-valued *N3D* spherical harmonics, [1],

$$p(\boldsymbol{r}, k) = \sum_m i^m j_m(kr) \sum_{n,\sigma} Y^{\sigma}_{mn}(\theta, \delta) B^{\sigma}_{mn}(k) \,, \qquad (1)$$

where

$$Y^{\sigma}_{mn}(\theta, \delta) = \sqrt{2m+1}\, \tilde{P}_{mn}(\sin\delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \end{cases} \qquad (2)$$

$$\tilde{P}_{mn}(\sin\delta) = \sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}\, P_{mn}(\sin\delta)\ . \qquad (3)$$

For $n = 0$, σ only takes the value $+1$. θ here measures the angle around the coordinate symmetry axis, and $\pi/2 - \delta$ is the angle between the axis and the coordinate direction, so that δ would normally be called the elevation, as shown in Figure 2.

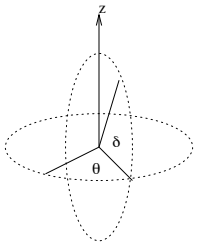

Figure 2: Spherical coordinates used.

From here on we shall use a slightly simplified notation that removes the need for σ by extending n to negative values as used in the standard complex set,

$$Y_{mn} = \begin{cases} Y^{+1}_{mn} & \text{if } n \geq 0 \\ Y^{-1}_{m|n|} & \text{if } n < 0 \end{cases}\ . \qquad (4)$$

Similarly the encoded signals become $B_{mn}(k)$.

2.2. Conversion to binaural

In a binaural rendering system the listener is presented with one signal to each ear canal direct. Binaural signals can be derived from an Ambisonic encoding, using Head Related Transfer Functions (HRTFs),[6], as described below. The Ambisonic encoding is easily rotated, which facilitates compensation for head movement.

Conversion to binaural can be achieved approximately by summing speaker array feeds that are each filtered by an HRTF matching the speaker position, [7, 8]. Figure 3 illustrates the signal flow in this process.

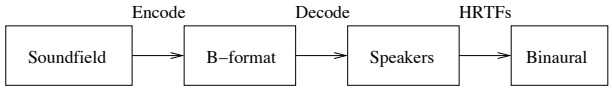

Figure 3: Encoding a soundfield to binaural via virtual speakers.

A natural extension of this idea to an exact method for binaural signals, is to transform the encoded soundfield into a planewave expansion, and weight each component planewave by the planewave HRTF matching its direction and frequency, [9, 10]. The process can be applied to high-order Ambisonic encodings containing sources at various distances. A straightforward binaural approach would require HRTF sets for each source distance, however decoding the high-order signal requires only the planewave HRTF set. This is not too surprising, as the HRTF sets are defined within the constraints of the wave equation, and so are all related. There is another less obvious advantage, which is that complex sources can be conveniently converted to binaural via an encoding in high-order B-format, as will be demonstrated later in this article. A single nearfield HRTF set cannot be applied in a simple way to a complex source description to yield the required binaural signals. Figure 4 depicts an overview of the encoding process from soundfield to binaural using a planewave expansion.

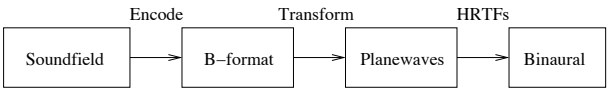

Figure 4: Encoding a soundfield to binaural via a planewave expansion.

The process is exact in the farfield, but as explained later, there is a subtle source of error which can affect near sources. First we detail the steps to generate binaural signals from high-order B-format using a planewave expansion.

We aim to represent a source-free region by an expansion in planewaves, known as a Herglotz expansion, with coefficients $\mu(\boldsymbol{s}, k)$ defined over unit vectors s, so that

$$p(\boldsymbol{r}, k) = \frac{1}{4\pi}\int_{S_u} dS(\boldsymbol{s})\, e^{ik\boldsymbol{s}\cdot\boldsymbol{r}}\mu(\boldsymbol{s}, k)\ , \qquad (5)$$

where integration is over the unit sphere. The spherical harmonic expansion of the source-free region, using standard complex spherical harmonics Y^n_m corresponding to N3D harmonics Y_{mn}, is

$$p(\boldsymbol{r}, k) = \sum_m j_m(kr)\sum_n Y^n_m(\theta,\delta)A^n_m(k)\ . \qquad (6)$$

$j_m(kr)$ are the spherical bessel functions. From Eqs. (6) and (5) valid planewave coefficients can be found in terms of the spherical harmonic coefficients A^n_m [10],

$$\mu(\boldsymbol{s}, k) = \sum_{m,n} i^{-n} A^n_m(k) Y^n_m(\boldsymbol{s})\ , \qquad (7)$$

and in terms of the N3D convention,

$$\mu(\boldsymbol{s}, k) = \sum_{m,n} B_{mn}(k) Y_{mn}(\boldsymbol{s})\ . \qquad (8)$$

The lack of a complex factor in Eq. (8) reflects the fact that in Ambisonics, plane waves with zero phase at the center have real-valued encodings, allowing the identification to be made between microphone polar patterns and the N3D harmonics. From the linear supposition of planewaves, the binaural signals, $p^L(k), p^R(k)$ are found by integrating the planewave weights with HRTF responses, $H^L(k\boldsymbol{s}), H^R(k\boldsymbol{s})$, over the sphere,

$$p^L(k) = \int_{S_u} dS(\boldsymbol{s})\mu(\boldsymbol{s}, k)H^L(k\boldsymbol{s})\ , \qquad (9)$$

for the left side and similarly for the right. In practice the integral can be replaced by a quadrature sum, with very little loss of accuracy for a sufficient number of quadrature points, of order the number of spherical harmonics in (8), [10].

In addition, we can explicitly calculate the nearfield HRTF of wavenumber k and given position, by binaurally synthesizing a pure nearfield monopole with that k and position. This is given by $p^L(k)$ in Eq. (9). Figure 5 summarizes this. The planewave expansions of monopoles are investigated further in the next section.

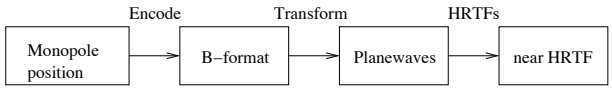

Figure 5: Finding a nearfield HRTF from a monopole encoding.

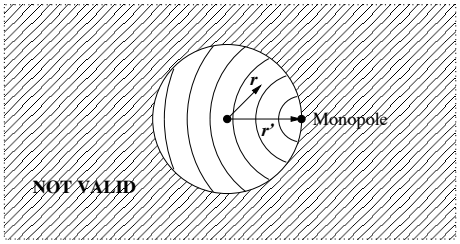

Figure 6: Spherical harmonic freefield expansion of a displaced monopole.

3. SCATTERING OF NEARFIELD SOURCES

3.1. Spherical expansion of a monopole

To study expansions of near sources we look in detail at the monopole. The important features are also true for general sources. A monopole source at non-central position $\boldsymbol{r'}$ has the following expansion in $\boldsymbol{r}$, valid only for $r < r'$, [11],

$$\frac{e^{-ik|r-r'|}}{|r-r'|} = ik \sum_{m=0}^{\infty} j_m(kr) h_m(kr') \sum_{n=-m}^{m} Y_{mn}(\theta', \delta') Y_{mn}(\theta, \delta) \ , \tag{10}$$

where $j_m(kr')$ and $h_m(kr)$ are the spherical Bessel function of the first kind and spherical Hankel function of the second kind. The positive frequency convention is chosen which gives an outward moving wave for a time piece $e^{i\omega t}$. For the following discussion it is important to emphasize the region of validity for the expansion is a sphere extending as far as the source, as illustrated in Figure 6. For more general sources that extend over a region the valid region extends as far as the maximum radius that does not enclose part of any source.

A freefield expansion cannot represent a monopole outside the valid region because it must retain zero divergence everywhere. In the region of the monopole's origin the expansion flows inwards to the freefield origin. The bessel functions $j_m(kr)$ are very close to zero for $kr < m$, so for a general freefield expansion $m_{max} \approx kr'$ is sufficient to synthesize accurately in the region $r < r'$. A detailed error analysis has been given, [10].

3.2. Scattering validity of synthesized sources

An HRTF filter generates the signal at an ear resulting from the scattering of a plane wave by the listener. The phase of the planewave is assumed to be zeroed to the center of the head. We can express the resultant field as a sum of the original unscattered field and the scattered component, $p = p_{in} + p_{scat}$. The scattering can be formulated in terms of the Sommerfield radiation conditions, which state that p_{scat} depends only on p_{in} at the boundary of the scattering body. As we have seen, the Ambisonic representation of a nearfield source is accurate only within a limited region, no matter how high the order of approximation. The derived planewave decomposition, can also only be accurate within the limited region. The example in Figure 7 is formed from 196 plane waves, whose directions are distributed around the sphere on Fleige nodes, [12]. One symmetric half of the normalized field is shown, $Re(p)/|p|$. If part of the scattering body is outside this region, then p_{in} is no longer correct on all of the scattering body. The binaural signals, found according to Eq. (9), are part of the resultant field p, and so in general suffer loss of accuracy when the valid region does not enclose the scattering surface. Figure 8 illustrates this for a front view of a listener, where the source is close enough that everything from the shoulders downwards is excluded from the valid region. It has been shown previously that the torso plays a significant role in

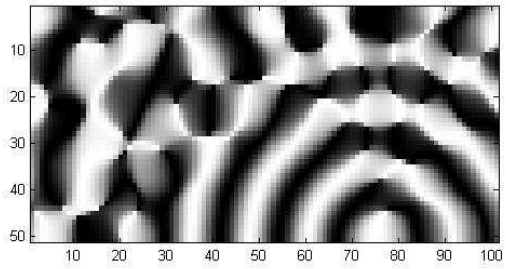

Figure 7: Planewave-from-spherical expansion of a monopole with $r' = 2\lambda, m_{max} = 12$, 196 Fliege nodes.

The arrows in Figure 8 show the flow of energy in the freefield expansion of the source. In a field with a real source, both arrows would point away from the source. It is evident that the scattering in the shoulder region using the freefield will be quite different to the scattering with a source field, and cause differences in the resultant field at the ears.

3.3. Higher accuracy nearfield expansions

The above result is not too surprising in retrospect, because we should not expect to be able to construct the response from a source embedded in an arbitrarily complex scattering geometry using only planewave scattering responses. However, the question remains, how well can we do with planewave HRTFs? Planewave expansions have a useful property that spherical harmonic expansions lack, they can be translated and expressed about a different point simply by multiplying by phase factors: If a position relative to the new center is $\boldsymbol{r'}$, and the corresponding position relative to the

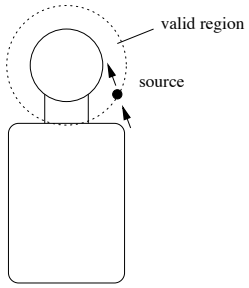

Figure 8: Parts of scattering body outside the valid expansion region of a near source.

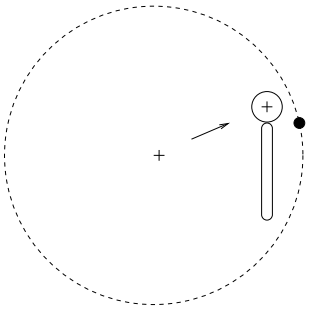

Figure 9: Shifted planewave expansion center for full scattering with a near source.

old center is $\boldsymbol{r}$, then $\boldsymbol{r} = \boldsymbol{r}' + \boldsymbol{x}$, where $\boldsymbol{x}$ is the translation from the old to the new center. So $e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = e^{i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{r}')} = e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}'}$. Therefore from Eq. (5) the expansion coefficients about the new center are $\mu'(\boldsymbol{k}) = e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\mu(\boldsymbol{k})$. Using this result, we can take the expansion for a source at a large radius, then shift the center so the source is at the required relative location. In the process the region of validity has been expanded, so a greater scattering body can be included, and the resulting binaural signals will be more accurate. Figure 9 illustrates this, showing a body entirely within the valid region for scattering, and with a source near to the head.

There is a price to be paid for the improved rendering of the source however. The new region of validity excludes other near sources, which must be rendered separately with their own shifts, rather than from a single harmonic expansion. Also the new spherical expansion must be specified to higher order, because the region being scattered is at a greater radius in the freefield expansion. As a result of this the number of planewaves and hence HRTFs must be increased. The number of nodes required is $O(kr)$ where r is the radius of curvature of the region boundary. However, on the positive side, the corrected planewave expansions can be summed to create a single planewave expansion for several near sources. This would suggest that planewave expansions would be a more efficient representation of sources. However they can only be accurate in up to half of the space, and so must be transformed according to their position relative to the listener.

The limit of the boundary of validity obtained by shifting, is a plane through the source, so it is impossible, as conjectured earlier, to precisely generate binaural signals, using planewave HRTFs, for sources in concave regions of the scattering body, such as under the chin.

Figures 10 and 11 show two plots of the magnitude $|\mu(\phi)|$ of the planewave decomposition of a freefield expansion of a monopole source at $r = 2\lambda, 4\lambda$ respectively. ϕ is the spherical coordinate measuring the angle between the direction and the coordinate symmetry axis, with $\phi = 0$ being the direction to the source. The decomposition of a source in a general position is just a rotation of $\mu(\phi)$. The greater detail seen in Figure 11 compared with Figure 10, reflects that it is sensitive to higher resolution HRTF data.

4. TRANSFORMING COMPLEX SOURCE ENCODINGS TO A FREEFIELD ENCODING

As discussed in the introduction, we wish to generalize the previous discussion on binaural rendering from monopole sources to general complex sources. First we consider an appropriate representation for the source, and then the transformation to a freefield

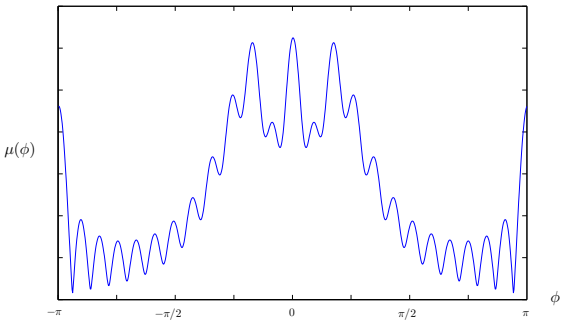

Figure 10: Planewave expansion coefficient for $r = 2\lambda$, $m_{max} = 12$.

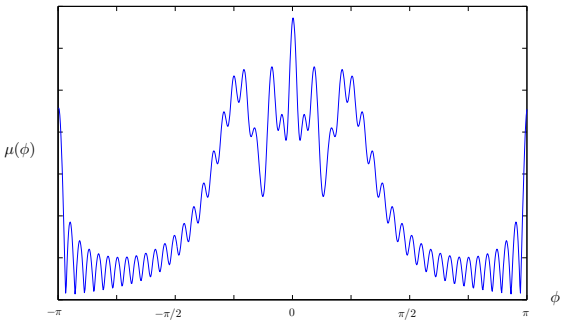

Figure 11: Planewave expansion coefficient for $r = 4\lambda$, $m_{max} = 24$.

encoding that can be applied to the binaural rendering technique of the last section.

4.1. The exterior harmonic expansion

A natural way to represent a general source region is the *exterior expansion* for the wave equation. This is valid outside a radius containing all parts of the source. The basis functions in the frequency domain using spherical coordinates are $h_m(kr)Y_{mn}(\theta, \delta)$, where $h_m(kr)$ are the spherical hankel functions of the second kind, [11]. m is the multipole order of each function, and $k = 2\pi/\lambda$ is the wavenumber. The type of hankel function chosen gives an outward moving wave when associated with a positive frequency time piece $e^{i\omega t}$, the same convention used in [2].

An infinitesimally defined multipole of order m can always be expressed exactly using an exterior expansion with terms up to order m. For this reason an exterior expansion is alternatively called an *exterior multipole expansion* or just a *multipole*, [14]. Another term used is *singular expansion*, since the center of the expansion has a singularity. The exterior expansion relates closely to the non-uniform directivity of a source, as discussed below, and our goal shall be to manipulate it to provide an Ambisonic source encoding. By multipole we shall mean an exterior expansion, unless otherwise stated.

The remainder of this section reviews the exterior expansion and introduces the conventions that will be used. We again adopt the N3D convention defined by (2). For convenience we define

coefficients, $O_{mn}(k)$, by a general exterior expansion,

$$p(\boldsymbol{r},k) = k\sum_m i^{-m-1} h_m(kr) \sum_n Y_{mn}(\theta,\delta) O_{mn}(k) \ , \quad (11)$$

so that in the farfield where $h_m(kr)$ tends to $i^{m+1}e^{-ikr}/kr$, the field becomes

$$p_{far} = \frac{e^{-ikr}}{r} \sum_{m,n} Y_{mn}(\theta,\delta) O_{mn}(k) \ . \quad (12)$$

The $O_{mn}(k)$ coefficients then directly express the non-uniform directivity in this regime, where locally the field tends to an outward moving plane wave. In the the $O_{mn}(k)$ coincide with the *O-format* encoding used previously for Ambisonic synthesis, [15, 9]. The same name will be used here for the more general case described by (11). We emphasize that this is just a convention, for convenience and appropriate to its context, in the same sense as B-format is defined.

$O_{mn}(k)$ can be readily calculated from measurements of the field on a sphere at any radius r outside the source region. Applying an integral over the sphere, $\int d\Omega\, Y_{mn}(\theta,\delta)$ to (11) gives

$$O_{mn}(k) = \frac{i^{m+1} \int d\Omega\, Y_{mn}(\theta,\delta) p(\boldsymbol{r},k)}{4\pi k h_m(kr)} \ . \quad (13)$$

For a real object the field could be measured approximately with pressure microphones placed located on a sphere a fixed distance from the source. In the farfield where the field becomes planar, inwardly pointing directional mics are equally effective given the appropriate equalization. For devices such as loudspeakers that convert electricity to sound linearly, the process can be simplified by repeated response measurements with a single mic that is moved. Speaker simulations might for instance be useful in high-end architectural simulations. When the $O_{mn}(k)$ responses are convolved with input signals for the speakers, the expansion signals are generated.

4.2. Source approximation order and error

We consider now the order to which a source is approximated, m_{max}. We wish to minimize this subject to reconstruction error constraints. A source can be arbitrarily small and still have power up to any multipole order, for example using the explicit definition of infinitesimal multipoles. However this is unusual in a real acoustic source because opposed component sources are not usually found very close together. Detailed error analysis, [14], shows that for a more typical source, the relative error decreases rapidly through 1 at $m_{max} \approx 2\pi r'/\lambda = kr'$, where r' is the radius of the source.

4.3. Multi-resolution sources

So far we have considered using single multipole expansions. In some cases a hybrid approach may be more appropriate, in which a source is represented using a several multipoles. This is neccessary whenever we wish to find the field at a free space inside the bounding sphere of an object, for example nearer to a table surface than its length. Outside the total bounding sphere a single multipole is sufficient. As we move closer to some part of the source more multipoles are neccessary. Far from the object compared to its size the field can be approximated as a plane wave, using the O-format coefficients to determine the direction dependence. This scheme of successive simplification resembles the multi-resolution techniques common in computer graphics, [16].

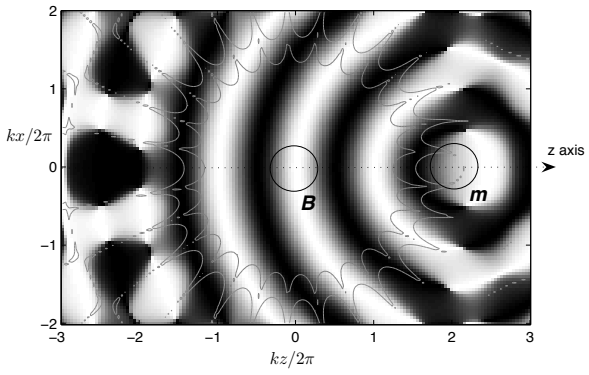

Figure 12: Cross-section of a field plot for a 13th order freefield expansion, center at **B**, of a monopole, center **O**. Error contours are shown at the 1% and 10% levels. The cross-section is $\theta = 0$. x, z are cartesian coordinates in length units.

5. AMBISONIC ENCODING OF MULTIPOLES

As we have seen, high-order Ambisonics is founded on the freefield or interior expansion, (1). Consider a field containing a single monopole set away from the freefield expansion center. Eq. (10) can be recast as the freefield expansion for a monopole, by fixing $\boldsymbol{r}$ and instead varying $\boldsymbol{r}'$. The form of this expansion is then consistent with Eq. (1), from which the values of the $B_{mn}(k)$ can be read off, as shown in [2]. The condition of convergence $\boldsymbol{r}' < \boldsymbol{r}$ now implies that the expansion converges *within* a circle that just touches the monopole source.

Figure 12 shows a normalized field plot $Re(p)/|p|$ for such a monopole reconstructed to the 13th order, and set at a distance 2λ from the expansion center. Outside this area the expansion is a valid freefield, although no longer matches the source field. Overall convergence behaviour within the valid region is like any other freefield, although close to the monopole, $\delta < \lambda$, the order required to achieve a given error is increased compared to a smooth freefield, as we would expect, [14]. The limit set to the region of freefield convergence by the source can not be exceeded by increasing the freefield order.

Higher multipole sources must also have freefield expansions about a displace origin, since they can be expressed as a sum of infinitesimal monopoles. The freefield expansion is then valid within a radius that does not include any of the source.

5.1. Multipole to freefield coefficient transformation

The main task in this section is to find $B_{mn}(k)$ in the presence of a multipole described by $O_{mn}(k)$ at a given position. It would be desirable to find a generalized closed form expression, as for the monopole case in [2]. However, it is not very apparent how this could be done or even if it would be the most practical method of calculation. Instead a more pragmatic approach is adopted yielding eventually a manageable integral expression. To begin (1) and (11) are equated. The notation is modified according to Figure 13,

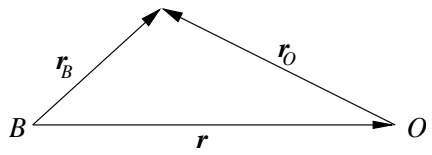

Figure 13: Vector notation

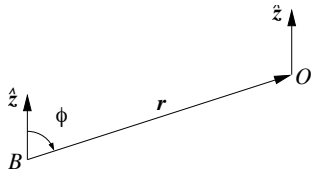

Figure 14: Finding components relative to $\boldsymbol{r}$

$$\sum_m i^m j_m(kr_B) \sum_n Y_{mn}(\theta_B, \delta_B) B_{mn}(k) = k \sum_m i^{-m-1} h_m(kr_O) \sum_n Y_{mn}(\theta_O, \delta_O) O_{mn}(k) \tag{14}$$

To isolate $B_{mn}(k)$ the operator $\int d\Omega_B \; Y_{m'n'}(\theta_B, \delta_B)$ is applied, with r_B a freely chosen constant, and θ_O, δ_O and r_O are functions of the vector $\boldsymbol{r}_B$, yielding

$$4\pi i^{m'} j_{m'}(kr_B) B_{m'n'}(k) = \sum_m i^{-m-1} \sum_n O_{mn}(k) \int d\Omega_B \; Y_{m'n'}(\theta_B, \delta_B) Y_{mn}(\theta_O, \delta_O) h_m(kr_O). \tag{15}$$

Relabeling indices, $B_{mn}(k)$ can be written as

$$B_{mn}(k) = \sum_{m',n'} M_{mnm'n'}(k, \boldsymbol{r}) O_{m'n'}(k), \tag{16}$$

where the filter matrix $M_{mnm'n'}(k, \boldsymbol{r})$ is

$$M_{mnm'n'}(k, \boldsymbol{r}) = \frac{k i^{-m-m'-1}}{4\pi j_m(kr_B)} \int d\Omega_B \; Y_{mn}(\theta_B, \delta_B) Y_{m'n'}(\theta_O, \delta_O) h_{m'}(kr_O) \; . \tag{17}$$

The $\boldsymbol{r}$ direction dependence in the matrix can be factored out by transforming the components $B_{mn}(k)$ and $O_{mn}(k)$ so they are relative to $\boldsymbol{r}$. Figure 14 shows the relationship between the initial coordinate axis, $\hat{\boldsymbol{z}}$, and the vector, $\boldsymbol{r}$, connecting the centres B and O. $\phi = \pi/2 - \delta$ together with θ specify a rotation mapping $\hat{\boldsymbol{z}}$ onto $\boldsymbol{r}$, written in components as $R_{m'n'n}(\theta, \phi)$. The third degree of freedom is unspecified, although it must be consistent. Therefore $R_{m'n'n}(\theta, -\phi)$ transforms $B_{mn}(k)$ and $O_{mn}(k)$ to find their coordinates relative to $\boldsymbol{r}$,

$$B'_{m'n'}(k) = \sum_n R_{m'n'n}(\theta, -\phi) B_{m'n}(k) \tag{18}$$

$$O'_{m'n'}(k) = \sum_n R_{m'n'n}(\theta, -\phi) O_{m'n}(k) \tag{19}$$

Now Eq. (16) can be written with an $\boldsymbol{r}$-direction-independent matrix, $M_{mnm'n'}(k, r)$,

$$B'_{mn}(k) = \sum_{m',n'} M_{mnm'n'}(k, r) O'_{m'n'}(k), \tag{20}$$

where

$$M_{mnm'n'}(k, r) = \frac{k i^{-m-m'-1}}{4\pi j_m(kr_B)} \int d\Omega_B \; Y_{mn}(\theta_B, \delta_B) Y_{m'n'}(\theta_O, \delta_O) h_{m'}(kr_O). \tag{21}$$

The coordinates in the integral are now relative to $\boldsymbol{r}$, although they haven't been relabeled. The symmetry this brings, with r_O independent of $\theta_B = \theta_O$, can be used to factor the integral into a product of θ and δ integrals. To make this clear $Y_{mn}(\theta, \delta)$ is factored into

$$Y_{mn}(\theta, \delta) = \hat{P}_{mn}(\sin\delta) \times \begin{cases} \sqrt{2}\cos n\theta & \text{if } n > 0 \\ 1 & \text{if } n = 0 \\ \sqrt{2}\sin n\theta & \text{if } n < 0 \end{cases} \tag{22}$$

where for convenience later, $\hat{P}_{mn}$ is defined,

$$\hat{P}_{mn}(\sin\delta) = \sqrt{(2m+1)\frac{(m-|n|)!}{(m+|n|)!}} \; P_{m|n|}(\sin\delta) \tag{23}$$

and $P_{mn}(x)$ is the associated Legendre polynomial. (21) becomes

$$M_{mnm'n'}(k, r) = \frac{k i^{-m-m'-1}}{4\pi j_m(kr_B)} \int d\delta_B \; \cos\delta_B \hat{P}_{mn}(\sin\delta_B) \hat{P}_{m'n'}(\sin\delta_O) h_{m'}(kr_O) \times 2\pi\delta_{nn'} = \frac{\delta_{nn'} k i^{-m-m'-1}}{2 j_m(kr_B)} \int_{-1}^{+1} ds_B \hat{P}_{mn}(s_B) \hat{P}_{m'n'}(s_O) h_{m'}(kr_O) \; , \tag{24}$$

where $s_B = \sin\delta_B$ and $s_O = \sin\delta_O$. s_O and r_O can be found from r, r_B and s_B using $r_B s_B - r_O s_O = r$. $r_O = r\sqrt{1 + \alpha^2 - 2\alpha s_B}$ and $s_O = r(\alpha s_B - 1)/r_O$, where $\alpha = r_B/r$. Setting $\alpha = .51111$ ensures good numerical behaviour.

Eq. (20) can now be simplified using,

$$B'_{mn}(k) = \sum_{m,n} \frac{1}{r} M_{mnm'}(kr) O'_{m'n}(k), \tag{25}$$

where we define a simplified matrix coefficient with 3 indices rather than 4.

$$M_{mnm'}(k) = \frac{ki^{-m-m'-1}}{2j_m(r_B)} \int_{-1}^{+1} ds_B \hat{P}_{mn}(s_B)\hat{P}_{m'n}(s_O)h_{m'}(r_O), \tag{26}$$

with $r = k$ implicit. Clearly this is defined only for $n < m$ and $n < m'$. This means for a given source the number of filters increases only linearly with B-format order required. The new filter coefficients are given in terms of one parameter, k. The actual filter acting in (25) is scaled in frequency by the radius r and there is a distance factor $1/r$. Putting this together with (18) gives $B_{mn}(k)$ in terms of $O_{mn}(k)$,

$$B_{mn}(k) = \sum_{n'} R_{mnn'}(\theta,\phi) \sum_{m'} \frac{1}{r} M_{mn'm'}(kr) \sum_{n''} R_{m'n'n''}(\theta,-\phi)\, O_{m'n''}(k) \tag{27}$$

Note that an orientation rotation could be incorporated into the first rotation acting on O_{mn}.

5.2. Validation and properties

To provide an immediate confidence test that the derived formulas are correct, a random test 5th order multipole was constructed, shown in Figure 15, and compared with the 13th order freefield expansion calculated using the matrix (27), shown in Figure 16. The error contours in Figure 16 at 10% and 1% levels are for deviations from the original multipole shown in Figure 15. The region of agreement extends as far as the center of the original multipole, as expected, and supports the derivations in this section.

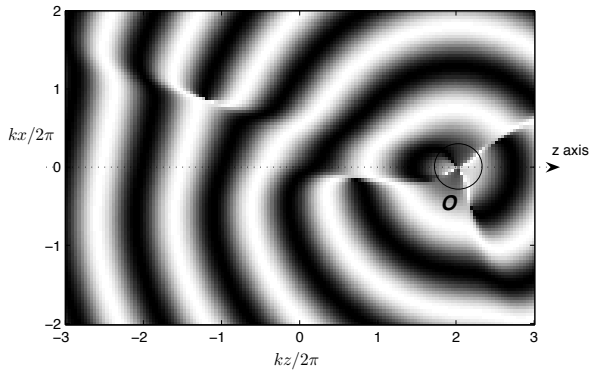

Figure 15: Cross-section of a field plot for a 5th order multipole, center at **O**. The cross-section is $\theta = 0$. x, z are cartesian coordinates in length units.

Next we examine $M_{mnm'}(kr)$ by checking that it is consistent with previous results for the monopole case,[2], in which the encoded signal is given by $B_{mn} = S(k)F_m(kr)Y_{mn}(\theta,\delta)$, where $F_m(kr) = i^{-m}h_m(kr)/h_0(kr)$. To match the alignment used to define $M_{mnm'}(kr)$, $\theta = \delta = \pi/2$. We first note that the source term $S(k)$ includes the delay and distance attenuation so that $S(k) = \frac{e^{ikr}}{r}O_{00}(k)$. In order to isolate the part matching

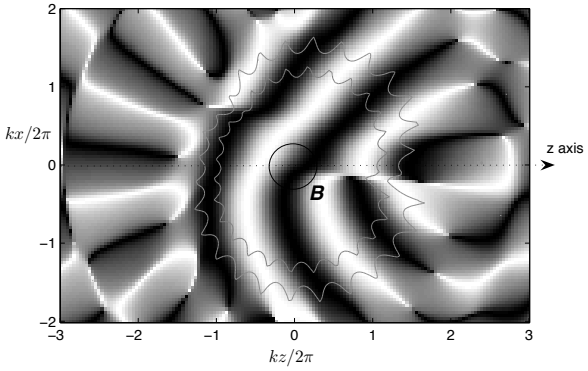

Figure 16: Cross-section of a field plot for a 13th order freefield expansion, center at **B**, of a multipole, center **O**. Error contours are shown at the 1% and 10% levels. The cross-section is $\theta = 0$. x, z are cartesian coordinates in length units.

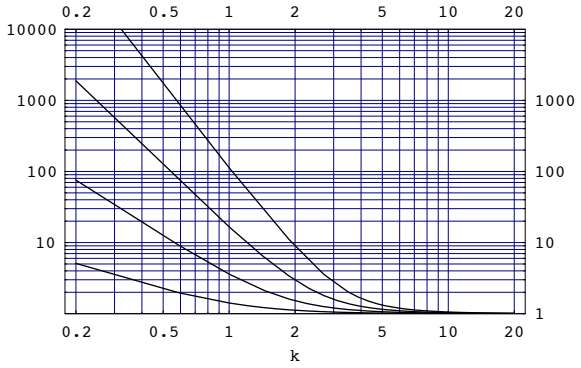

Figure 17: Amplitude response for $M_{m00}(kr)/(e^{-ikr}/r)/Y_{m0}(0,\pi/2), m = 1,2,3,4$

$F_m(kr)$, we look at the adjusted value,

$$M_{m00}(kr)/(e^{-ikr}/r)/Y_{m0}(0,\pi/2)\ .$$

With $r = 1$, this produces the plots shown in Figure 17, matching previous results, [2]. The general picture at higher multipoles is that, with the e^{-ikr}/r piece factored out, the response is always minimum phase. For small k the order of the filter becomes $m + m'$, while for large k it is n. The transition occurs around $k = 2$ corresponding to ≈ 0.3 wavelength separation from the source. $M_{mnm'}(kr)$ has symmetries which reduce the computational cost of using it; $M_{mnm'}(kr) = M_{m'nm}(kr)(-1)^{m+m'} = M_{m-nm'}(kr)$. The first of these is useful for cross checking the accuracy of value, since the integrals are different.

In [2] filters were commuted from the decoding stage to control the large amplitudes at low values of k. The situation at first appears worse here because filters encoding high multipoles can have much higher order for the same order m of the B-format encoding. However, at the values of k where the filters grow large, the size of the source object in wavelengths becomes small, and a lower source order suffices to approximate it well, according to

the discussion in Section 3 where a regular source is considered. This means that low frequencies can be filtered out progressively from the higher orders of object encoding, so avoiding excessive gains. The filtering should be performed dynamically according to distance. The highest order retained, m_{max} is determined by the highest frequency, f_{max}, that is required to be accurately reconstructed, as described in Section 3. For an accurately constructed field of sufficiently low frequency, a monopole suffices.

We do not investigate the implementation of the filter matrix in detail here, but note that in general it will take a similar form to that described for the monopole case, [2]. Because the filters are evaluated numerically, they must be converted to IIR form by pole/zero fitting. Filter modification according to radius can then achieved by scaling the poles and zeros in frequency.

5.3. Reverberation encoding and transformation

This article has focused on the synthesis of the direct signal from complex objects in the nearfield. Related techniques can be applied to the synthesis of a reverberant signal originating from a complex source. There is not space here to give the full details, so we only outline the results. First, we can encode the reverberant response to complex source, for a given pair of source and listener positions using a filter matrix $M^{rev}_{mnm'n'}$ which has been measured from a real acoustic or calculated using ray tracing or similar methods.

$$B_{mn}(k) = \sum_{m',n'} M^{rev}_{mnm'n'}(k, \boldsymbol{x}_B, \boldsymbol{x}_O) O_{m'n'}(k) \ , \qquad (28)$$

Secondly, we can transform the reverberant field, B_{mn}, to the field based at another listener location, B'_{mn}, using a variant of the matrix in (26),

$$B'_{mn}(k) = \sum_{n'} R_{mnn'}(\theta, \phi) \sum_{m'} \frac{1}{r} M^{BB}_{mn'm'}(kr) \sum_{n''} R_{m'n'n''}(\theta, -\phi)\, B_{m'n''}(k) \ , \qquad (29)$$

where

$$M^{BB}_{mnm'}(k) = \frac{k i^{m'-m}}{2 j_m(r'_B)} \int_{-1}^{+1} ds_B \hat{P}_{mn}(s'_B) \hat{P}_{m'n}(s_B) j_{m'}(r_B) \ , \qquad (30)$$

The advantage of these methods is that we can efficiently generate high quality reverberation, with controllable listener position and source orientation.

6. CONCLUSION

The paper began by looking at how nearfield sources can be rendered binaurally using only planewave HRTFs. An inherent source of error was found owing to the nature of the freefield harmonic expansion. This can be corrected at the expense of being able to use a single harmonic expansion to encode all nearfield sources. It is possible however to sum the corrected planewave expansions to give a single planewave expansion for all near sources that exist outside the convex hull of the scattering envelope of the listener.

The paper continued by deriving a transformation law from a multipole source to a freefield harmonic expansion at a point outside the source. This can then be converted to a corrected planewave expansion and used to render binaural signals valid for nearfield locations. The processing of reverberant signals is also mentioned using related techniques.

We have not presented any tests applying our methods with real HRTF data. This is clearly an important and complex task, which we hope to address. Working with high quality personalized HRTFs will be a key factor. With the on-going refinement of virtual reality systems, the considerations presented here are expected to become increasingly relevant.

7. REFERENCES

[1] J. Daniel, *Reprsentation de Champs Acoustiques, Application la Transmission et la Reproduction de Scnes Sonores Complexes dans un Contexte Multimdia*, Ph.D. thesis, University of Paris 6, Paris, France, 2000.

[2] J. Daniel, "Spatial sound encoding including near field effect," in *Proc. AES 23nd International Conference*, 2003.

[3] D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," *Journal of the Acoustical Society of America*, vol. in press, 2007.

[4] M.A. Gerzon, ," *Journal of the Audio Engineering Society*, vol. 33, pp. 859–871, 1985.

[5] M.A. Gerzon, "General metatheory of auditory localisation," in *Proceedings of the 92nd AES Convention*, 1992.

[6] C. Roads, *The Computer Music Tutorial*, 2002.

[7] A. McKeag and D. McGrath, "Sound field format to binaural decoder with head tracking," in *Preprint 4302, AES Convention 6r*, 1996.

[8] J.M. Jot and S. Wardle, "Approaches to binaural synthesis," in *Proceedings of the AES 105th Convention*, 1998.

[9] D. Menzies, "W-panning and o-format, tools for object spatialisation," in *Proc. AES 22nd International Conference*, 2002.

[10] R. Duraiswami, D.N. Zotkin, Z Li, E. Grassi, N.A. Gumerov, and L.S. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with hrtf cues," in *Proceedings of the AES 119th Convention*, 2005.

[11] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*, 1968.

[12] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal on Numerical Analysis*, vol. 19.2, pp. 317–334, 1999.

[13] V.R Algazi, C. C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *Journal of the Acoustical Society of America*, vol. 109, pp. 11101122, 2001.

[14] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*, Elsevier Science, 2005.

[15] D. Menzies, *New Performance Instruments for Electroacoustic Music*, Ph.D. thesis, University of York, UK, 1999.

[16] M. S. Dodgson, N. A.and Floater and M. A. (Eds.) Sabin, *Advances in Multiresolution for Geometric Modelling*, Springer, 2005.